Introduction to the Varenius Project
M. Goodchild, M. Egenhofer, K. Kemp, D. Mark, E. Sheppard
International Journal of Geographical Information Science 13 (8): 731-745, 1999

Introduction to the Varenius Project

Michael F. Goodchild[1]
Max J. Egenhofer[2]
Karen K. Kemp[3]
David M. Mark[4]
Eric Sheppard[5]

Abstract

This paper introduces a special issue of the journal on the subject of Project Varenius, a three-year effort funded by the US National Science Foundation to advance geographic information science. Geographic information is first defined as an abstraction of primitive tuples linking geographic locations to general descriptors. Geographic concepts originate in the human mind, and are instantiated in geographic information. Geographic information technologies apply digital methods to geographic information. Finally, geographic information science is defined as the set of basic research issues arising from these technologies. Three motivations are presented for research in this area: scientific, technological, and societal. Within the project, geographic information science is structured by a three-part framework that includes cognitive, computational, and societal issues. The paper ends with an introduction to these three parts, which define the infrastructure of the project and are discussed at length by the subsequent three papers.

1. Introduction

The US National Science Foundation, the nation's largest sponsor of basic research, is currently funding a substantial effort, known as Project Varenius, to advance the science of geographic information. Support for the project began in February 1997 and will extend into February 2000; it is being provided through a cooperative agreement with the site of the National Center for Geographic Information and Analysis at the University of California, Santa Barbara. This introductory paper to this special issue describes the background to the project, its intellectual and scientific objectives, its motivation, and its structure. Three major papers follow on each of the strategic areas of Varenius research.

Although this journal changed part of its name in 1996 from Systems to Science, and although the term *geographic information science* (GIScience) seems to be catching on in various ways, as evidenced by the growth of the US-based University Consortium for Geographic Information Science (*http://www.ucgis.org*), nevertheless there are lingering uncertainties about exactly what

---

[1] National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA. Phone: 805 893 8049. FAX: 805 893 3146. Email: good@ncgia.ucsb.edu.
[2] National Center for Geographic Information and Analysis, and Department of Spatial Information Science and Engineering, University of Maine, Orono, ME 04469-5711, USA.
[3] Geographic Information Science Center, University of California, Berkeley, CA 94720, USA.
[4] National Center for Geographic Information and Analysis, and Department of Geography, University at Buffalo, Box 610023, Buffalo, NY 14261-0023, USA.
[5] Department of Geography, University of Minnesota, Minneapolis, MN 55455, USA.

geographic information science means, about its relationship to geographic information technologies, and about how to move it forward. The Varenius project proposal developed a narrow and possibly provocative view of the field, and the first part of this introduction describes this view, and the reasons that led to its adoption. It is followed by a discussion of the possible components of the field, and the three strategic areas of the project are introduced. The final section briefly describes the history and structure of the project, and the reasons for naming it in honor of a 17th Century Dutch geographer; and leads into the three major papers that form the largest part of this special issue.

2. Geographic information

Geographic information (GI) can be defined as information about the features and phenomena located in the vicinity of the surface of the Earth. What distinguishes this particular type of information from other types is of course the presence of a reference to some geographic location, and all GI can be reduced eventually to a simple statement that at some location there exists an instance of some more generally recognized *thing,* where *thing* might be a class, a feature, a concept, a measurement of some variable, an activity, an organism, or any of a myriad possibilities. By geographic location we mean to include up to three spatial variables, depending on whether distance above or below the surface is important, and also time, if the thing is time-dependent. More formally, the fundamental primitive element of geographic information is the tuple $<x,y,z,t,\mathrm{U}>$ where U represents some thing present at some location (x,y,z,t) in space–time.

Many authors have reviewed the nature of the geographic space–time in which geographic information is embedded (Gatrell 1983, Couclelis 1998). The dimensions of the Earth and the accuracy of surveying technology limit the range of reasonable resolutions of geographic space, from perhaps 10^7 m at the high end, or the linear dimension of the planet as a whole, to 1 m, which is the finest resolution associated with commonly-available Earth imagery and mapping, or perhaps as small as 1 mm, the limit of most practical surveying. Within these limits there are no complications from relativistic or quantum effects, and no reason not to treat geographic space–time as a simple, rigid Newtonian frame[6]. Within that frame the geographic surface is a two-dimensional curved object, and it is generally convenient to redefine the spatial coordinates as latitude, longitude, and elevation (ϕ,λ,h), based on some agreed mathematical approximation to the true form of the surface; unfortunately, this leaves our primary method of defining location dependent on which of many vertical and horizontal datums is chosen for the purpose. Moreover, it was also convenient in the era when most GI had to be recorded on flat sheets of paper to transform latitude and longitude to a two-dimensional frame (u,v) using one of a large number of transformation functions or map projections $P(\phi,\lambda)$ (Bugayevskiy and Snyder 1995).

We conceive of the geographic frame as continuous, so a complete characterization of some variable such as temperature at all points in space, or in space–time, leads to an infinite number of tuples. GI is manageable therefore only under certain conditions which render the number of recorded tuples finite (in what follows the symbol $\boldsymbol{x}$ refers to some location; the number of space–time dimensions defining $\boldsymbol{x}$ will be determined by the context):

- Correlation. This allows aspects of what is present at $\boldsymbol{x}$ to be guessed from other aspects of what is present at $\boldsymbol{x}$. We humans have become adept at identifying only a minimal number of

---

[6] However, the ways in which human and physical systems respond within that frame are complex, and many authors have been led to adopt a non-Newtonian frame in attempting to understand them.

things at a location, and inferring other things from them.

- Spatio–temporal autocorrelation. This allows aspects of what is present at $\boldsymbol{x}$ to be guessed, predicted, or inferred from what is present at some other location $\boldsymbol{x}+\boldsymbol{e}$. In general the ability to predict declines in both space and time with the magnitude of $\boldsymbol{e}$, since spatial autocorrelation is usually stronger over short distances (Tobler's first law of geography; Tobler 1970), and many geographic phenomena also vary slowly if at all through time.

- Things of finite spatial extent. If a thing is recorded as present over a region (a connected set of locations) rather than at a single location, then a single tuple of GI can connect the thing with the region. In this way a large number of tuples of the form $<x,y,z,t,\mathrm{U}>$ can be replaced by a single tuple <R,U> where $(x,y,z,t)\in \mathrm{R}$. This condition can be generalized to include things that form a recognizable pattern over the region, where the pattern is not necessarily the simplest one of uniformity. For example, the function $f(\boldsymbol{x}) = a + bx + cy$; $\{\boldsymbol{x}\in \mathrm{R}\}$ implies a pattern of linear spatial variation over region R.

- Named regions. A set of locations R may be sufficiently well-known and stable to be worth naming. Provided everyone can agree on the tuple <N,R>, which identifies a place-name N with a region R, then tuples of the form <N,U>, describing a thing present in the region named U, are an economical way of recording GI. Unfortunately there is often lack of agreement over the exact geographic extent named N. *Gazetteers*, which are a major source of relations of tuples <N,R>, all too often approximate R by a single central point (ϕ,λ).

- Generalization. The number of tuples can be made manageable if the thing recorded at a location is allowed to deviate from the thing that is actually observed. If *most* things within a region are equal, it may be acceptable to record a single thing for the entire region. If the thing has numerical value, it may be acceptable to record only a mean for a region, rather than values at every location within the region. If the spatial variation of a numerical thing over the region is closely matched by a mathematical function, it may be acceptable to record only the function. A vast range of possible generalization rules are in common use in the domain of GI, both to record information and also to derive simpler approximations from information that is already recorded. The *differences* between what is recorded and what is observed (or believed to be true) are the basis for the study of uncertainty in GI (see, for example, UCGIS 1996: 125).

- Abstraction. If we can agree on a standard vocabulary of things, then the larger the vocabulary, the greater the prospects for information reduction. For example, the abstracted tuple <R,*lake*> can be expanded first into a large number of tuples <$\boldsymbol{x}$,*lake*> for all $\boldsymbol{x}\in \mathrm{R}$, and expanded further based on our understanding of the nature of the thing *lake* (e.g., uniformly fresh water, uniform elevation). The expansion of <R,*hill*> is more problematic because *hill* is a much less specific indication of the things present at each of the locations in R.

The set of possible abstractions is vast, although limited in some domains by attempts to establish authorities or thesauri of standard geographic terms (see, for example, the work of the U.S. Federal Geographic Data Committee, *http://www.fgdc.gov*, and Rugg *et al*. 1997), and discourage the use of others. It includes abstractions that connect things in one location at one time to things in other locations at other times, through conceptualizations of dynamic processes that affect geographic landscapes. It includes *geographic prepositions*, which define instances of relationships between more primitive things, as in *the road crosses the park*, or *the house is next* *to the road* (e.g., Mark and Egenhofer 1994).

If information is to be shared between people, it is essential that its meaning be understood by all of the parties to the sharing. In scientific practice, information is expected to satisfy the requirement of replicability, that any two observers of the same phenomenon would record the same information, to within the accuracy of their measuring instruments. It follows that the same two observers would interpret the information in the same way. In the scientific world the bases of measurement are subject to international agreement and standardization, and are free of ambiguity.

In the case of geographic information, the system of spatio–temporal referencing defined by (ϕ,λ,h,t) is defined according to international standards that have been agreed since the late 19th Century (when the Greenwich Meridian was accepted as the global standard of longitude), although problems exist because of the multiplicity of datum standards in use. However some ambiguity still persists because of imperfect measuring instruments, which prevent two observers agreeing exactly on the spatio–temporal location of a given point. Two observers using simple GPS receivers to record the location of a point might disagree by as much as 100m. But much larger *semantic* problems exist because of lack of agreement on the meaning of *things*, the other part of the primitive geographic tuple (for recent discussions of semantic issues in GI see Kuhn 1997, Harvey 1998). Some things, notably physical variables, can be defined sufficiently well to meet the scientific standard. Others are *fuzzy* (e.g., Burrough and Frank 1996), meaning that there is sufficient ambiguity in their definition to lead to greater variation among observers than is attributable to inaccuracies in measuring instruments. Others are subject to regional variations in dialect, or variations in terminology between disciplines or social classes. Some, such as *hill*, are so far away from the norms of scientific description as to be virtually useless except in the loosest description.

Despite this range, however, we state as a basic principle that *GIScience must embrace the study of GI that fails to meet the scientific criterion*, although GIScience itself must be based on rigorous scientific principles. Although much of the information that is collected and processed by GI technologies is scientific, it is also true that GI technologies are widely used in other contexts, and that much of our knowledge of the world fails to meet rigid scientific standards of observation. It is also true that ambiguity is not always bad, and that precision is not always better, especially when it is inconsistent with accuracy. GI technologies may be asked to answer questions that are ambiguous but nevertheless important to the originator, such as 'Is Santa Barbara north of Los Angeles?' They may be used to process data, from sources such as soil maps, that include information that is inherently fuzzy, but nevertheless is believed to be useful in certain contexts.

Perhaps the most powerful argument for this basic principle rests on expectations about the nature and interests of general GIS users. People habitually work with GI in ways that fall short of scientific standards of rigor, just as they use other precise technologies like the telephone to communicate imprecise information. They give each other directions, for example, that are full of poorly-defined and ambiguous terms, such as *near, not far*, *about*. A GIS that hopes to be useful in this environment can be designed according to one of two principles. First, it can be designed to work only along rigorous scientific lines, and require that users adjust their normal behaviors to match, insisting, for example, that they not use terms like *near* in interacting with the system, unless they have first given *near* rigorous definition. Alternatively, a GIS can be designed to support normal discourse (for an example of GIS interaction using the term *near* see Robinson *et al.* 1986). Of course one might argue that the second option is impossible as long as the computer is a machine with perfectly mechanistic behaviors; or that it is impossible in principle to design a precise machine that can interface smoothly with a less-than-perfect human intelligence. Nevertheless, recent experience in computing technologies does seem to support the notion that computing applications can be designed so as not to intrude significantly on the normal cognitive processes of their users. At worst, the possibility is an important research question.

3. Geographic concepts

In the previous section we argued that GI is constructed from primitive tuples <$\mathbf{x}$,U> through processes of abstraction, use of common vocabularies, and generalization. We use the term *geographic concepts* to describe all of the generic components of GI, including:

- the concepts that provide the basis for GI itself (e.g., the geodetic system, projections, metrics of distance);
- elements of geographic thesauri, the standard authorities that define a common vocabulary of feature types (e.g., *lake*, *reservoir*, *river*, *city*, *building*);
- the contents of gazetteers that establish the positions of named places;
- the terms and phrases that define relationships between geographic point sets (e.g., *near to*, *north of*, *crosses*, *intersects with*); and
- the generic classes of *things* that define the phenomena present at geographic locations (e.g., variables such as *elevation*, *temperature*; land cover classes; land ownership; zoning regulations).

Some of these concepts meet scientific standards of replicability, but others do not. Some are simple, such as metrics of distance, or systems of Earth coordinates, while others are far more complex, including *tropical storm*, *neighborhood* and *esker*. Some have meaning that is shared by virtually the entire human population, such as longitude, while others have meanings that are confined to disciplines, language groups, regions or other *information communities*.

4. Digital geographic information technologies

All of the previous discussion is independent in principle of whether digital computers are used in any stage of the handling of GI. In general, it is possible to represent geographic variation through models of two general classes: *analog* and *digital*. By an analog model, we mean that the model is represented in some space as a scaled replica of reality; the space is usually physical (e.g., a paper map), but may also be electrical (e.g., transmission of seismic records as analog electrical signals). By a digital model, we mean that the recorded properties are coded into some discrete alphabet, using an agreed set of rules. To all intents and purposes that alphabet is the binary alphabet, and *digital* implies the use of modern digital computer technology. Devising rules for recording important properties of the real world in the binary alphabet of the digital computer is one of the most challenging tasks of GIScience.

The use of digital technology conveys enormous advantages over analog modeling, and these have driven the rapid development and adoption of digital GI technologies over the past four decades. The relevant technologies include GPS, remote sensing, image processing, soft photogrammetry, the surveyor's total station, scanners, virtual environments, and plotters, all of which are to some degree specific to GI. Digital technology allows for easy editing, since there is no need to interact with a physical model; calculation and manipulation of data through the use of arithmetic and logical operations; reliable storage and handling, since it is much easier to protect digital systems from additional error and unwanted noise; and sharing, since digital information can be transmitted at the speed of light and at very low cost.

We use the term *geographic information system* (GIS) to describe the most generic and powerful of these technologies, embracing all forms of digital analysis, manipulation, querying, communication, retrieval, and output. Early versions of GIS were software applications on large mainframes; later, mainframes were replaced by desktop workstations. Today, however, GIS must be understood as a complex system of distributed data and processing resources, designed to support manipulation, analysis, modeling, and decision-making based on digital GI.

Digital information technologies have become so pervasive in today's world that it is difficult to find examples of information that is not digital at some stage in its life. Telephone voice communication has shifted almost transparently from analog to predominantly digital, and is moving rapidly to *packet-switching* technologies that handle conversations as small, independent packets of bits. The contents of paper sheets are now commonly transmitted using the digital protocols of FAX. Paper maps can be scanned, transmitted by FAX, and digitized using one of a large number of accepted rules. The mere existence of digital coding at some point in the life of the data is not very significant, however; much more significant are the methods used to code the data, and the constraints that this coding imposes. Conversion to digital form imposes constraints if its inverse fails to restore the full content of the information (by full content we mean all of the elements of the data that are known to be useful). For example, scanning the contents of a map with a resolution S (followed by its inverse, in this case plotting onto paper) deletes all of the variation at resolutions less than S; digitizing a smooth curve as a polyline (a sequence of points connected by straight lines) deletes any information not captured by the polyline.

There is nothing *inherent* in the information loss that results from coding in digital form, since with sufficient care it is possible to express virtually any information in digital form without loss. Communication is mediated by the senses, and we have abundant ways of representing visual and acoustic signals in digital form; only in the case of tactile and olfactory communication is there any suggestion that digital encoding may fail to capture signals. However, the constraints imposed by the coding schemes in common use for digital data almost always result in information loss, because they are normally chosen to achieve a balance between loss on the one hand, and volume of data and ease of manipulation on the other.

5. Geographic information science

We are now in a position to define GIScience. Information science generally can be defined as the systematic study according to scientific principles of the nature and properties of information. From this position it is easy to define GIScience as the subset of information science that is about GI.

This definition is straightforward, but it fails to address a key question: why does GI form a subset that requires specialized study, or what is special about GI? Several authors have addressed this question, normally under the rubric 'what is special about spatial?' (unfortunately there seems to be no acceptable equivalent for 'geographical'). Anselin (1989) takes a statistical approach, arguing that GI differs from the mainstream of statistical information in the almost ubiquitous presence of spatial dependence and spatial heterogeneity. Each of these properties deviates from the normal assumptions of statistical methods; specifically, spatial dependence causes problems for the independence assumption that underlies many tests, and spatial heterogeneity causes problems for assumptions of homoscedasticity and the stationarity of other statistical properties.

Rhind (1996) takes a broader perspective that includes the institutional and societal context of GI, arguing that economic, legal, and public policy issues define the special nature of GIS. We believe these arguments and those of Anselin are convincing, and fully justify GIScience (and see also Goodchild, 1992) as a separate and significant subset of information science.

An independent series of arguments leads to a justification for GIScience based on the needs of the GI technologies for basic research (Goodchild 1992; Wright *et al.*, 1997). GI technologies are so important and influential, it is argued, and raise such interesting basic questions, that a substantial investment is needed in the research issues that underlie the technology and determine its long-term development. Although large investments have been made in these technologies over the past three decades, there remain many impediments to greater efficiency, more effective analysis and modeling, and greater use; and these impediments may yield to sustained research in the appropriate disciplines.

Thus we have two distinct but convergent bases for defining GIScience, one deriving from information science generally, and the other from the GI technologies. Other terms may be equally acceptable, including *geomatics*, which is in common use in some parts of the research community; *spatial information science*, *geoinformatics*, and perhaps others.

GIScience re-examines some of the most fundamental themes in traditional spatially-oriented fields such as geography, cartography, and geodesy, in the context provided by the emerging digital age and the society in which it is embedded and which it influences. It incorporates recent developments in cognitive and information science, together with more specialized research in established disciplines such as computer science, statistics, mathematics, and psychology. It also incorporates developments in our understanding of the nature of society, and the forces that structure and influence it. It is motivated in part by questions of why certain geographic problems cannot easily be addressed with current GIS software. Similarly, the difficulty that some people have in using state-of-the-art GISs, or in moving from one GIS to another, raises basic questions about human spatial cognition, about how to capture and represent geographic knowledge in an information system, and about the continuing interference by immature technology in the performance of substantive tasks.

On a more technical level, GIScience needs to resolve a number of questions such as:

- How do we *conceptualize* geographic worlds?
- How are geographic worlds *constructed* by society, and how does society affect our understanding of geographic space?
- How do we *capture* (measure) geographic concepts, recognize them in the field or in remotely-sensed information, and identify their accuracy and quality?
- How do we *represent* geographic concepts, especially in digital environments, with

incomplete information, with alternative data models in the same systems, and perhaps with different representations of the same phenomena (multiple representation)?

- How do we *store, access*, and *transform* geographic concepts with as little loss of information as possible during data sharing, file transfer, and data archiving?
- How do we *explain* geographic phenomena through the application of appropriate methods of analysis, and models of physical and human processes.
- How can geographic information technologies help to *reveal* and *constrain* understandings and interpretations, particularly in the human sciences?
- How do we *visualize* geographic concepts, as two-dimensional static maps on printed media or electronic displays, or as animated displays or hypertext documents?
- How do we *use* geographic concepts to think about geographic phenomena, to make decisions about places, and to seek explanations for geographic patterns and phenomena?

Traditionally, such questions have been addressed by researchers working within existing disciplines, and much progress has been achieved. However, the work has been spread across many research fields, and often has been conducted within very different research traditions. No systematic conceptual framework has emerged from these relatively isolated efforts. Commonalties among the questions and their solutions may be missed in fragmented research environments. Furthermore, cross-disciplinary work is often risky in academia, especially to early-career researchers, because promotion standards often given highest priority to scientific outlets that are defined by the boundaries of traditional disciplines, and tend to assign greatest prestige to their centers. We believe that by addressing these questions within the framework provided by the emerging field of GIScience, we can help to reduce institutional impediments to progress in these research areas, and to encourage the exploration of issues in ways that go beyond the solution of immediate problems.

6. Motivations for geographic information science

In addition to pure intellectual curiosity, we see research in GIScience as motivated from three distinct directions: *scientific, technological*, and *societal*.

6.1 Scientific motivation

Research in GIScience serves the needs of science and scientists in two ways. First, it addresses areas where our understanding of key geographic notions and their appropriate representations is currently incomplete. We see such basic research as especially important at this time in areas where human conceptualizations and digital implementations of concepts interact and conflict. Second, GIScience contributes to the conceptualizations, methods, and tools with which scientists approach geographically-distributed phenomena. Thus it contributes to the infrastructure of science, particularly for those disciplines whose subject matter is distributed over the Earth's surface, and for which a geographic perspective is likely to prove useful.

Historically, the development of the GI technologies has been influenced only in a limited way by the needs of science. GISs have their roots in government agency data-gathering and decision-making, the design disciplines like landscape architecture and planning, and the mapping sciences of cartography and surveying. GPS was a military development. Only in remote sensing has there been a longstanding link to scientific applications. But this situation has changed rapidly in the past decade, particularly in disciplines like anthropology, hydrology, and terrestrial ecology, where the broadly-based functionality of GIS is able to provide a comprehensive software environment for a terrestrially-based science. Not surprisingly, much interest in GIS has originated in geography, with its comprehensive interest in phenomena on the surface of the Earth. In other fields, however, such as oceanography (Wright and Goodchild, 1997) and atmospheric science, the role of GIS is currently limited to preprocessing of boundary conditions and visualization, since the two-dimensional, static map metaphor used in current GISs has proven too restrictive for sciences concerned with the transient behavior of fluids in three dimensions. On the other hand certain generic issues of GIScience, such as error modeling and propagation in spatial data, are eminently relevant to these sciences also.

As in many other instances, the development and adoption of GIS tools in the scientific community raises questions about the influence of tools on the conduct of science, and whether such tools can alter the ways of doing science as the microscope and the telescope did in the past (Abler, 1987). Should the scientist insist on knowing exactly what operations are performed by the tools, or is this principle bound to be weakened as science becomes more complex, more collaborative, and more interdisciplinary? Do the choices that the use of a geographic database imposes on its users constrain the science that can be performed, in ways that are often out of the immediate control of the scientist? Under what circumstances is the scientist willing to trust data that he or she did not collect, and will the increased technological ability to share scientific data over the Internet and using the World-Wide Web (WWW) change them (Onsrud and Rushton, 1995)? Such questions about tools often have their roots in theoretical questions about appropriate representations, operations, and concepts.

6.2 Technological motivation

Our second motivation derives from the technology push that we are currently experiencing. The proliferation of faster computing hardware and the emergence of an information infrastructure are quickly changing the ways people think and work, creating digital worlds.

> 'While the politicans struggle with the baggage of history, a new generation is emerging from the digital landscape free of many of the old prejudices. These kids are released from the limitation of geographic proximity as the sole basis of friendship, collaboration, play, and neighborhood.' (Negroponte, 1995 p. 230).

New information technologies have significant influence on the advancement of GIScience through the design and use of GISs. They enable geoscientists to collaborate in new ways, sharing large spatial data collections or performing tasks together without the need to be present at the same location at the same time. Digital worlds embed new paradigms. They move bits rather than atoms, offering access to data and information without any need to relocate physical media. They offer everyone the chance to make information publicly available. The telephone has been the medium for the exchange of voice, but the new information highways allow users to collaborate by exchanging digitally-coded data that stand for text, voice, images, and more complex structures such as GI.

Digital worlds form a new culture of computing, in which the user is paramount.

> 'It is important in focusing on what's ahead in communications, to zero in not on the technology, but what we use technology for. No one says "Let's use the telephone." They say "Let's call Grandma."' (Gore, 1993)

Ease of use is critical, and only the user's tasks should matter; systems should reflect the needs of users, without requiring them to be concerned with technicalities. The current limitations of GIS are such that it is clear that the technology will have to be reinvented repeatedly, and we doubt whether the GIS of ten years from now will be recognizable to its current practitioners. Of particular importance are visual representations, which replace numeric representations, and visual thinking. While visual representation boosts communication ('A picture is worth a thousand words'), it hides internal representations. This perception causes a dilemma for scientists who in general desire to understand fully what the system they are using does with their data. One might argue that the cultures of science and digital worlds diverge on this issue, counter to the commonly-held belief that computing provides an ideally supportive environment for science.

While many commentators present a uniformly rosy picture of a technology-based future, the academic sector has a useful role to play in dispassionate assessment. Geographic information science should also be about the flaws in bullish predictions, including questions of equity, narrowness of representation, and many other issues. These are addressed at length in this volume by Sheppard *et al.*

6.3 Societal motivation

Bullish predictions of a glowing future for digital technology and GIS aside, we also believe it is important that such a fast-growing and groundbreaking technology be subjected to the kinds of dispassionate reflection at which the academic sector excels. No other group is likely to take the kinds of long, hard looks at GIS that are needed if its benefits to society are to be maximized, and its potential abuses avoided or controlled.

Geographic perspectives are fundamental to an understanding of the interplay between local and global environments; the couplings between physical processes in the terrestrial subsurface, atmosphere and oceans, and their interactions with the human world; and the integration of processes and policies over geographically varying boundary conditions. As such, they may offer approaches to the solution of many of society's most pressing problems. For example, the US National Research Council report *Science Priorities for the Human Dimensions of Global Change* concluded that:

> 'substantial advances have been made in GISs, which allow the merging of population data with other data using geographic location as a join point, (and) GISs allow the population research community to bring its considerable statistical, methodological, and theoretical skills to issues that heretofore have not been researchable.' (NRC, 1994)

While we can claim only the most indirect of linkages between GIScience and the solution of problems of global hunger, unemployment, or crime, each of these issues provides a context in which GI technologies play an important part, and where it is important that these tools be as carefully thought out and as effective as possible.

At a more immediate level, GIScience research addresses such issues as the implications of GI technologies for systems of democratic representation; the potential for popular empowerment through concepts of electronic democracy; the legal liability associated with GI and GI technologies; the potential for invasion of privacy, surveillance, and control; and the implications of GI technologies for the organization of human activities in geographic space. Many similar issues arise in connection with all aspects of digital technology; an important question for GIScience in each case is whether the geographic context makes the problem in any sense unique.

7. The components of geographic information science

Based on these definitions and motivations, we can now begin to examine GIScience in detail, and to discuss various ways of partitioning the field. Our guiding principle in doing so is the need for research progress: how best can we move the field forward?

Goodchild (1992) reviewed the various disciplines that might have a role in contributing to progress in GIScience. These included the disciplines that have traditionally focused on GI, and on one or more of the associated technologies. Cartography largely predates the digital era, as do surveying, geodesy, and photogrammetry, though all four have been enormously stimulated by it. The GIScience disciplines also should include those that have substantial contributions to make, but for which GI has not been seen as an important focus in the past: in this category we include statistics (spatial statistics), economics (information economics); cognitive science (spatial cognition), psychology (environmental and developmental psychology, and social psychology), and mathematics (geometry).Geography and computer science clearly have contributions to make to GIScience, though in both cases the significance of the field within the broader objectives of the discipline is open to debate (in the case of geography, see Wright *et al.* 1997). We believe, however, that any partitioning of the field of GIScience based on traditional disciplines will work against the needs of the field, which are surely best addressed through multidisciplinary collaboration.

The US-based University Consortium for Geographic Information Science (UCGIS) has taken a consensus-building approach, asking each of its institutional members and their delegates to identify key research areas, and distilling a structure from the response. At its 1996 Assembly the UCGIS arrived at ten topics (Table 1), and these were further refined at the 1998 Assembly. Papers on each of the topics are available at *http://www.ucgis.org*, and a summary paper has been published (UCGIS 1996). The ten topics are roughly equal, however, only in the level of support they received from the membership; no other coherent basis has been suggested for dividing the field in this way.

Table 1: The ten research topics of the US-based University Consortium for Geographic Information Science (source: UCGIS 1996)

Acquisition
Uncertainty
Spatial analysis
Future of the information infrastructure
Interoperability
Distributed and mobile computing
GIS and society
Scale

Cognition
Representations

The Varenius project uses a method of structuring the field that differs from both of these previous options in having a coherent intellectual basis. We propose that the domain of GIScience addresses three distinct arenas:

- the individual, as user of technology, observer of geographic phenomena, source of conceptualizations, and maker of decisions;
- the system, defined as the entire complex of digital GI technologies and its supporting hardware, software, and networking foundations; and
- society, including its institutions, customs, communities, norms, and standards.

GIScience must address specific questions in all three arenas, including:

Individual: how do people conceptualize the world around them, and reason about it using those conceptualizations?

System: how can we design GIS to achieve maximum performance and functionality, with minimum information loss or other constraint?

Society: what processes determine the adoption of GIS in society, and its use by institutions, and how does the adoption of GIS change the way society constructs space?

But more significant perhaps are the questions that arise at the boundaries between the three arenas, such as:

Individual–System: to what degree does the use of digital coding constrain the ability of individuals to record and communicate knowledge of the geographic world?

System–Society: what will be the impact on societal issues, such as privacy, as a result of rapid growth in the use of GIS, and how is GIS a construction of society?

Society–Individual: how will GI technologies change the individual's access to information, and the ability of governments to monitor society's members?

This tripartite division also allows us to identify the areas of GIScience most in need of attention by specific disciplines, to encourage participation by members of those disciplines in research, and to develop research ideas within a coherent framework. Below we identify the disciplines we believe are most likely to contribute to each of the three research areas:

Individual: cognitive science, environmental psychology, linguistics;

System: computer science, information science;

Society: economics, sociology, social psychology, geography, political science.

More specifically, the following titles seem to us to summarize the important issues of geographic information science:

Cognitive Models of Geographic Space (issues of the individual);

Computational Methods for Representing Geographic Concepts (issues of the system);

Geographies of the Information Society (issues of society).

We have struggled with an appropriate title for the societal area, and the title given is far from satisfactory, but we are unable to find a better one. *Geographies of the Information Society* suggests an undue emphasis on future patterns of human activity, under the influence of such technologies as telecommuting, and thus misses all of the other issues that arise because of the complex interactions between GI technologies and society. However, we continue to use it as the title of this area (note that Sheppard *et al.* also refer to this area as the more anonymous 'Apex').

8. The Varenius project

The project is designed around this tripartite structure. Each of the three *strategic areas* is overseen by a panel of international experts, with the responsibility to monitor progress and manage project activities in the area. The three panels are chaired by David Mark (Cognitive Models of Geographic Space), Max Egenhofer (Computational Methods for Representing Geographic Concepts), and Eric Sheppard (Geographies of the Information Society). During the lifetime of the project they will each sponsor workshops on significant and promising research topics in the area, and undertake related activities to foster progress. The three papers that form the largest part of this special issue describe the three research areas and their activities in detail.

The project is named for Bernardus Varenius, who is perhaps best known as the author of the *Geographia Generalis* (1650), a work that ranks among the most influential in the history of the discipline of geography. The influence of Varenius on his contemporaries, particularly Sir Isaac Newton, has been documented by Warntz (1981, 1989) and will not be reviewed here. Rather, we cite two simple reasons for honoring him in this way:

- The division of the discipline of geography into *general* and *special* branches, rather than the more conventional *human* and *physical*. This division avoids the problems that arise when separate human and physical understanding must be integrated. More importantly, it identifies general geography as including the principles of Earth measurement, and the dynamic processes that create the landscape, while special geography documents the unique characteristics of places, and the boundary conditions of processes. This cleavage dominated methodological discussion in the discipline of geography for much of this century (Johnston 1991); but it finds a compelling resolution in GISs, which combine the special (the data) with the general (the algorithms, principles, concepts, and models) into a functioning whole. We make this argument advisedly, however, being aware of the limited ability of GIS to deal with dynamics, and thus with representation of process.
- The conviction, clearly apparent to Newton, that the geography of Varenius was founded on principles of scientific observation and reasoning.

The following three papers describe the three areas of the project in detail. Each presents a definition of its domain, a review of our current level of understanding, and a prospectus for the future. The three papers vary somewhat in approach, as appropriate to the nature of each topic.

The primary mechanism of the Varenius project is the specialist meeting, which brings together an international group of experts in a workshop setting to review progress in a given area, identify researchable topics that will move the area forward, and define a research agenda. Within the project, mechanisms exist for pursuing topics after the specialist meeting through seed grants for proposal preparation, and funds to support visits to other institutions to develop collaborations. Each workshop also results in a report, which is published in the NCGIA Varenius series and made available electronically. Further details on the Varenius project can be found at the NCGIA web site, *http://www.ncgia.org*. They include information on past and future meetings, meeting reports, and additional information on the project structure and administration.

The project includes resources to support three specialist meetings for each of the three strategic research areas. The topics for these were chosen by the respective panels, as representing areas of high significance to geographic information science where rapid progress might be anticipated. Further details of each, and the basis on which they were chosen, are given in the three papers. The papers also include results of the meetings and subsequent research where these were available at the time of writing.

9. Acknowledgment

The authors gratefully acknowledge the support of the U.S. National Science Foundation, which provides core support for the Varenius project under Cooperative Agreement SBR 9600465 to the University of California, Santa Barbara.

10. References

Abler, R. F., 1989, The National Science Foundation National Center for Geographic Information and Analysis. *International Journal of Geographical Information Systems*, **1**(4), 303–326.

Anselin, L., 1989, *What is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis*. Technical Report 89-4 (Santa Barbara: National Center for Geographic Information and Analysis).

Bugayevskiy L. M., and Snyder, J. P., 1995, *Map Projections: A Reference Manual* (London: Taylor and Francis).

Burrough, P. A., and Frank, A. U., 1996, *Geographic Objects with Indeterminate Boundaries* (London: Taylor and Francis).

Couclelis, H., 1998, Space, time, geography. In P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind (editors) *Geographical Information Systems: Principles, Techniques, Management and Applications* (New York: Wiley).

Gatrell, A. C., 1984, *Distance and Space: A Geographical Perspective* (New York: Oxford University Press).

Goodchild, M. F., 1992, Geographical information science. *International Journal of Geographical Information Systems*, **6**(1), 31–47.

Gore, A., 1993, Remarks at the National Press Club.

Harvey, F., 1998, Designing for interoperability: overcoming semantic differences. In M. F. Goodchild, M. J. Egenhofer, C. A. Kottman, and R. Fegeas (editors) *Interoperating Geographic Information Systems* (New York: Kluwer).

Johnston, R. J., 1991, *Geography and Geographers: Anglo-American Human Geography since 1945*. 4th Edition (London: Edward Arnold).

Kuhn, W., 1997, Approaching the issue of information loss in geographic data transfers. *Geographical Systems*, **4**(3), 261–276.

Mark D. M., and Egenhofer, M. J., 1994, Modeling spatial relations between lines and regions: combining formal mathematical models and human subjects testing. *Cartography and Geographic Information Systems*, **21**(4), 195–212.

National Research Council, 1994, *Science Priorities for the Human Dimensions of Global Change* (Washington: National Research Council).

Negroponte, N., 1995, *Being Digital* (New York: Knopf).

Onsrud, H.J. and G. Rushton, editors, 1995, *Sharing Geographic Information* (New Brunswick: Center for Urban Policy Research).

Rhind, D. W., 1996, Economic, legal, and public policy issues influencing the creation, accessibility, and use of GIS databases. *Transactions in GIS*, **1**(1), 3–12.

Robinson, V. B., Blaze, M., and Thongs, D., 1986 Representation and acquisition of a natural language relation for spatial information retrieval. *Proceedings, Second International Symposium on Spatial Data Handling, Seattle*, pp. 472–487.

Rugg, R. D., Egenhofer, M. J., and Kuhn, W., 1997, Formalizing behavior of geographic feature types. *Geographical Systems*, **4**(2), 159–180.

Tobler, W. R., 1970, A computer movie simulating urban growth in the Detroit region. *Economic Geography*, (supplement) **46,** 234–240.

University Consortium for Geographic Information Science, 1996, Research priorities for geographic information science. *Cartography and Geographic Information Systems*, **23**(3), 115–127.

Varenius, B., 1650, *Geographia Generalis* (Amsterdam: Elsevir Press).

Warntz, W., 1981, *Geographia Generalis* and the earliest development of American academic geography. In B. W. Blouet (editor) *The Origins of Academic Geography in the United States* (Hamden: Archon Books).

Warntz, W., 1989, Newton, the Newtonians, and the *Geographia Generalis Varenii*. *Annals of the Association of American Geographers*, **79**(2), 165–191.

Wright, D. J., and Goodchild, M. F., 1997, Guest editorial: Data from the deep: implications for the GIS community. *International Journal of Geographical Information Science*, **11**(6), 523–528.

Wright, D. J, Goodchild, M. F., and Proctor, J. D., 1997, Demystifying the persistent ambiguity of GIS as 'tool' versus 'science'. *Annals of the Association of American Geographers*, **87**(2), 346–362.